

Matematikai statisztika

Programozó Matematikus szak részére

PAP GYULA

KLTE Matematikai és Informatikai Intézet

4010 Debrecen, Pf.12

papgy@math.klte.hu

1. Bevezetés

1.1. A matematikai statisztika célkitűzései

Adott egy **minta**, amely alapján következtetéseket szeretnénk levonni bizonyos valószínűségekkel, eloszlás– illetve sűrűségfüggvényekkel, paraméterekkel kapcsolatban.

Például:

Becslélmélet: ismeretlen paraméterek közelítése

pontbecslés: a mintában megállapított selejtarány alapján szeretnénk megbecsülni az egész sokaságban levő selejtarányt

intervallumbecslés: konfidencia–intervallum (azaz megbízhatósági intervallum) keresése, vagyis például olyan intervallum meghatározása a mintaátlag körül, melybe a várható érték nagy valószínűséggel beleesik

Hipotézisvizsgálat: döntés bizonyos hipotézisek (feltevések) elfogadásáról vagy elvetéséről; például a mintaátlagnak a várható értéktől való eltérése belefér-e még a szabvány-előírásokba?

illeszkedésvizsgálat: a minta milyen eloszlásból származik, azaz milyen eloszlással közelíthető jól? (Vagyis döntés arról, hogy az adott minta illeszkedik-e az adott eloszláshoz?)

homogenitásvizsgálat: döntés arról, hogy vajon két adott minta ugyanabból az eloszlásból származik-e?

függetlenségvizsgálat: döntés arról, hogy vajon két adott minta független-e?

1.2. Statisztikai alapfogalmak

1.2.1. Definíció. *(Statisztikai) minta alatt valamely $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn értelmezett független, azonos eloszlású valószínűségi változók $\xi_1, \xi_2, \dots, \xi_n$ véges sorozatát értjük. A minta **elemszáma** n , a minta **alapeloszlása** pedig a $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók közös eloszlása.*

1.2.2. Definíció. Statisztika (pontosabban statisztikai függvény) alatt a mintaelemek valamely $T(\xi_1, \xi_2, \dots, \xi_n)$ függvényét értjük, ahol $T : \mathbb{R}^n \rightarrow \mathbb{R}$ olyan függvény, hogy $T(\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi változó.

Megjegyezzük, hogy ehhez elegendő például az, hogy T folytonos legyen, hiszen ekkor tetszőleges $c \in \mathbb{R}$ esetén

$$\{\omega \in \Omega : T(\xi_1(\omega), \dots, \xi_n(\omega)) < c\} = \{\omega \in \Omega : (\xi_1(\omega), \dots, \xi_n(\omega)) \in T^{-1}(-\infty, c)\} \in \mathcal{A},$$

azaz esemény, ugyanis T folytonossága miatt $T^{-1}(-\infty, c) = \{x \in \mathbb{R}^n : T(x) \in (-\infty, c)\}$ nyílt halmaz, ezért előáll nyílt intervallumok megszámlálható uniójaként.

1.2.3. Definíció. Mintaátlag:

$$\bar{\xi}_n := \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n},$$

vagyis a mintaelemek számtani közepe.

A mintaátlag nyilván statisztika, hiszen $\bar{\xi}_n = T_n(\xi_1, \xi_2, \dots, \xi_n)$, ahol

$$T_n(x_1, x_2, \dots, x_n) = \frac{x_1 + x_2 + \dots + x_n}{n}$$

folytonos függvény.

1.2.4. Definíció. Tapasztalati (empirikus) szórásnégyzet:

$$s_n^2 := \frac{(\xi_1 - \bar{\xi}_n)^2 + (\xi_2 - \bar{\xi}_n)^2 + \dots + (\xi_n - \bar{\xi}_n)^2}{n} = \frac{1}{n} \sum_{i=1}^n \left(\xi_i - \frac{1}{n} \sum_{j=1}^n \xi_j \right)^2,$$

vagyis a mintaelemek átlagtól való négyzetes eltéréseinek átlaga.

A tapasztalati szórásnégyzet is statisztika, hiszen $s_n^2 = T_n(\xi_1, \xi_2, \dots, \xi_n)$, ahol

$$T_n(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2$$

folytonos függvény.

1.2.5. Lemma. (Steiner-formula) Tetszőleges $c \in \mathbb{R}$ esetén

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - c)^2 - \left(\frac{1}{n} \sum_{i=1}^n (\xi_i - c) \right)^2.$$

Speciálisan ($c = 0$ választással)

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - (\bar{\xi}_n)^2.$$

Bizonyítás.

$$\begin{aligned} s_n^2 &= \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2 = \frac{1}{n} \sum_{i=1}^n ((\xi_i - c) - (\bar{\xi}_n - c))^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\xi_i - c)^2 - 2 \frac{1}{n} (\bar{\xi}_n - c) \sum_{i=1}^n (\xi_i - c) + (\bar{\xi}_n - c)^2, \end{aligned}$$

ahol

$$\frac{1}{n} (\bar{\xi}_n - c) \sum_{i=1}^n (\xi_i - c) = (\bar{\xi}_n - c)^2 = \left(\frac{1}{n} \sum_{i=1}^n (\xi_i - c) \right)^2$$

□

1.2.6. Definíció. A $\xi_1, \xi_2, \dots, \xi_n$ mintából képezett **rendezett minta**:

$$\xi_{1:n}^* \leq \xi_{2:n}^* \leq \dots \leq \xi_{n:n}^*.$$

A rendezett minta elemei statisztikák, ugyanis például $\xi_{1:n}^* = T_n(\xi_1, \xi_2, \dots, \xi_n)$, ahol

$$T_n(x_1, x_2, \dots, x_n) = \min\{x_1, x_2, \dots, x_n\}$$

folytonos függvény. Megjegyezzük, hogy a $\xi_{1:n}^*, \xi_{2:n}^*, \dots, \xi_{n:n}^*$ valószínűségi változók nem függetlenek és nem azonos eloszlásúak.

1.2.7. Definíció. A $\xi_1, \xi_2, \dots, \xi_n$ mintából képezett **tapasztalati medián**:

$$\begin{cases} \xi_{m+1:n}^*, & \text{ha } n = 2m + 1 \text{ páratlan,} \\ (\xi_{m:n}^* + \xi_{m+1:n}^*)/2, & \text{ha } n = 2m \text{ páros.} \end{cases}$$

A $\xi_1, \xi_2, \dots, \xi_n$ minta **terjedelme**:

$$\xi_{n:n}^* - \xi_{1:n}^*.$$

1.2.8. Definíció. A $\xi_1, \xi_2, \dots, \xi_n$ mintából képezett **tapasztalati (empirikus) eloszlásfüggvény**: $\mathcal{F}_n : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$,

$$\mathcal{F}_n(x) = \mathcal{F}_n(x, \omega) := \sum_{i: \xi_i(\omega) \leq x} \frac{1}{n} = \begin{cases} 0, & \text{ha } x \leq \xi_{1:n}^*, \\ \frac{k}{n}, & \text{ha } \xi_{k:n}^* < x \leq \xi_{k+1:n}^*, \\ 1, & \text{ha } x > \xi_{n:n}^*. \end{cases}$$

Vagyis minden rögzített $x \in \mathbb{R}$ esetén $\mathcal{F}_n(x, \cdot) : \Omega \rightarrow \mathbb{R}$ valószínűségi változó, ezért statisztika, mégpedig a $\{\omega \in \Omega : \xi_i(\omega) < x\} \in \mathcal{A}$ esemény relatív gyakorisága. Továbbá minden rögzített $\omega \in \Omega$ esetén $\mathcal{F}_n(\cdot, \omega) : \mathbb{R} \rightarrow \mathbb{R}$ eloszlásfüggvény, mégpedig annak a diszkrét eloszlásnak az eloszlásfüggvénye, melynek a lehetséges értékei a mintaelemek $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)$ realizációi, és ezen értékeket egyaránt $1/n$ valószínűséggel veheti fel.

1.2.9. Tétel. (Glivenko) (A matematikai statisztika alaptétele) Legyenek $\xi_1, \xi_2, \dots$ független, azonos eloszlású valószínűségi változók közös F eloszlásfüggvénnyel. Ekkor

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\mathcal{F}_n(x) - F(x)| = 0 \right) = 1.$$

Tehát a tapasztalati eloszlásfüggvény egyenletesen közelíti az (elméleti) alapeloszlást, midőn az n mintaelemszám tart végtelenbe. Megjegyezzük, hogy minden rögzített $x \in \mathbb{R}$ esetén a nagy számok erős törvénye alapján 1 valószínűséggel teljesül $\mathcal{F}_n(x) \rightarrow \mathbb{P}(\xi_1 < x) = F(x)$ ha $n \rightarrow \infty$, ezért $\mathbb{P} \left(\lim_{n \rightarrow \infty} |\mathcal{F}_n(x) - F(x)| = 0 \right) = 1$.

1.2.10. Definíció. Legyen $r \in \mathbb{N}$. A $\xi_1, \xi_2, \dots, \xi_n$ mintából képezett **tapasztalati r -edik momentum**:

$$m_{r,n} := \frac{1}{n} \sum_{i=1}^n \xi_i^r.$$

Nyilván $\bar{\xi}_n = m_{1,n}$, és a Steiner-formula alapján $s_n^2 = m_{2,n} - m_{1,n}^2$. Továbbá $m_{r,n}$ éppen az $\mathcal{F}_n$ (véletlentől függő) eloszlásfüggvényhez tartozó r -edik momentum.

Hisztogram (oszlopdiaagram) úgy kapható, hogy vesszük a valós számegyenesnek egy $-\infty < x_1 < x_2 < \dots < x_k < \infty$ beosztását, és az egyes részintervallumokra olyan magasságú téglalapokat rajzolunk, ahány mintalemez az illető részintervallumba esik; vagyis egy olyan lépcsősfüggvényről van szó, amelynek értéke az egyes részintervallumokon az illető részintervallumba esés gyakorisága. **Sűrűségi hisztogram** úgy kapható, hogy a relatív gyakoriságokat ábrázoljuk. Ha a mintaelemszám végtelenbe tart és a beosztássorozat finomodik, akkor a sűrűségi hisztogram jól közelíti a sűrűségfüggvényt.

2. A statisztikában használatos néhány fontos eloszlás

2.1. χ_k^2 -eloszlás

2.1.1. Definíció. Ha $\eta_1, \eta_2, \dots, \eta_k$ független, standard normális eloszlású valószínűségi változók, akkor

$$\chi_k^2 := \eta_1^2 + \eta_2^2 + \dots + \eta_k^2$$

eloszlását **χ_k^2 -eloszlásnak** nevezzük, melynek **szabadsági foka k** .

2.1.2. Definíció. Szemifaktoriális:

$$n!! := \begin{cases} n(n-2)(n-4) \dots 1, & \text{ha } n \text{ páratlan,} \\ n(n-2)(n-4) \dots 2, & \text{ha } n \text{ páros.} \end{cases}$$

2.1.3. Lemma. A χ_k^2 -eloszlás abszolút folytonos, és sűrűségfüggvénye

$$f_{\chi_k^2}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ c_k x^{k/2-1} e^{-x/2}, & \text{ha } x > 0, \end{cases}$$

ahol

$$c_k = \begin{cases} \frac{1}{\sqrt{2\pi} \cdot (k-2)!!}, & \text{ha } k \text{ páratlan}, \\ \frac{1}{2 \cdot (k-2)!!}, & \text{ha } k \text{ páros}. \end{cases}$$

Speciálisan

$$f_{\chi_1^2}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & \text{ha } x > 0, \end{cases} \quad f_{\chi_2^2}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{1}{2} e^{-x/2}, & \text{ha } x > 0. \end{cases}$$

Tehát a χ_2^2 -eloszlás megegyezik az $1/2$ -paraméterű exponenciális eloszlással.

Bizonyítás. Nyilván χ_1^2 eloszlásfüggvénye

$$F_{\chi_1^2}(x) = P(\chi_1^2 < x) = P(\eta_1^2 < x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ P(|\eta_1| < \sqrt{x}), & \text{ha } x > 0. \end{cases}$$

Ha $x > 0$, akkor

$$P(|\eta_1| < \sqrt{x}) = P(-\sqrt{x} < \eta_1 < \sqrt{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-u^2/2} du = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-u^2/2} du = G(\sqrt{x}),$$

ahol

$$G(y) = \frac{2}{\sqrt{2\pi}} \int_0^y e^{-u^2/2} du.$$

Nyilván $G'(y) = \sqrt{2/\pi} e^{-y^2/2}$, ezért $F_{\chi_1^2}$ deriválható $(0, \infty)$ -en:

$$F'_{\chi_1^2}(x) = G'(\sqrt{x}) \frac{1}{2\sqrt{x}} = \sqrt{\frac{2}{\pi}} e^{-x/2} \frac{1}{2\sqrt{x}},$$

ezért χ_1^2 abszolút folytonos és sűrűségfüggvénye

$$f_{\chi_1^2}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & \text{ha } x > 0. \end{cases}$$

Továbbá χ_2^2 eloszlásfüggvénye

$$\begin{aligned} f_{\chi_2^2}(x) &= f_{\eta_1^2 + \eta_2^2}(x) = \int_{-\infty}^{\infty} f_{\eta_1^2}(y) f_{\eta_2^2}(x-y) dy \\ &= \int_{-\infty}^{\infty} f_{\chi_1^2}(y) f_{\chi_1^2}(x-y) dy = \begin{cases} 0, & \text{ha } x \leq 0, \\ \int_0^x \frac{1}{2\pi y} e^{-y/2} \frac{1}{2\pi(x-y)} e^{-(x-y)/2} dy, & \text{ha } x > 0. \end{cases} \end{aligned}$$

Tehát ha $x > 0$, akkor

$$f_{\chi_2^2}(x) = \frac{1}{2\pi} e^{-x/2} \int_0^x \frac{dy}{\sqrt{y(x-y)}} = \frac{1}{2\pi} e^{-x/2} \int_0^1 \frac{dz}{\sqrt{z(1-z)}} = c e^{-x/2},$$

ahol $y = xz$ helyettesítést hajtottunk végre. A c konstans abból határozható meg, hogy

$$1 = \int_{-\infty}^{\infty} f_{\chi_2^2}(x) dx = c \int_0^{\infty} e^{-x/2} dx = 2c.$$

Tehát végülis

$$f_{\chi_2^2}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{1}{2} e^{-x/2}, & \text{ha } x > 0. \end{cases}$$

Az általános eset teljes indukcióval bizonyítható. Ha feltesszük, hogy az állítás igaz k -ra, akkor

$$\begin{aligned} f_{\chi_{k+1}^2}(x) &= f_{\chi_k^2 + \eta_{k+1}^2}(x) = \int_{-\infty}^{\infty} f_{\chi_k^2}(y) f_{\eta_{k+1}^2}(x-y) dy = \int_{-\infty}^{\infty} f_{\chi_k^2}(y) f_{\chi_1^2}(x-y) dy \\ &= \begin{cases} 0, & \text{ha } x \leq 0, \\ \int_0^x c_k y^{k/2-1} e^{-y/2} c_1 (x-y)^{-1/2} e^{-(x-y)/2} dy, & \text{ha } x > 0, \end{cases} \end{aligned}$$

így ha $x > 0$, akkor

$$\begin{aligned} f_{\chi_{k+1}^2}(x) &= c_1 c_k e^{-x/2} \int_0^x y^{k/2-1} (x-y)^{-1/2} dy \\ &= c_1 c_k e^{-x/2} \int_0^1 (xz)^{k/2-1} [x(1-z)]^{-1/2} x dz = \tilde{c}_{k+1} x^{(k+1)/2-1} e^{-x/2}, \end{aligned}$$

ahol $\tilde{c}_{k+1}$ abból határozható meg, hogy

$$1 = \int_{-\infty}^{\infty} f_{\chi_{k+1}^2}(x) dx = \tilde{c}_{k+1} \int_0^{\infty} x^{(k+1)/2-1} e^{-x/2} dx.$$

Mivel

$$\begin{aligned} I_{k+1} &:= \int_0^{\infty} x^{(k+1)/2-1} e^{-x/2} dx \\ &= [x^{(k+1)/2-1} (-2) e^{-x/2} dx]_{x=0}^{x=\infty} + (k-1) \int_0^{\infty} x^{(k-1)/2-2} e^{-x/2} dx = (k-1) I_{k-1}, \end{aligned}$$

így

$$\tilde{c}_{k+1} = \frac{1}{I_{k+1}} = \frac{1}{(k-1)I_{k-1}} = \frac{\tilde{c}_{k-1}}{k-1},$$

amiből $c_1 = 1/\sqrt{2\pi}$ és $c_2 = 1/2$ figyelembevételével kapjuk az állítást. $\square$

2.2. $\mathbf{t}_k$ -eloszlás (Student-eloszlás)

2.2.1. Definíció. Ha η és χ_k^2 független valószínűségi változók standard normális, illetve χ_k^2 -eloszlással, akkor

$$\mathbf{t}_k := \frac{\eta}{\sqrt{\chi_k^2/k}}$$

eloszlását **$\mathbf{t}_k$ -eloszlásnak (Student-eloszlásnak)** nevezzük, melynek **szabadsági foka k** .

A $\mathbf{t}_k$ -eloszlás sűrűségfüggvényének meghatározásához szükségünk van a következő lemmára.

2.2.2. Lemma. Legyenek ξ és ζ abszolút folytonos, független valószínűségi változók f_ξ , illetve f_ζ sűrűségfüggvénnyel. Ekkor ξ/ζ abszolút folytonos, és sűrűségfüggvénye

$$f_{\xi/\zeta}(x) = \int_{-\infty}^{\infty} |v| f_\xi(xv) f_\zeta(v) dv.$$

Bizonyítás. Nyilván ξ/ζ eloszlásfüggvénye

$$\begin{aligned} F_{\xi/\zeta}(x) &= \mathbf{P}(\xi/\zeta < x) = \iint_{u/v < x} f_{\xi,\zeta}(u, v) du dv \\ &= \int_{-\infty}^0 \left(\int_{xv}^{\infty} f_\xi(u) f_\zeta(v) du \right) dv + \int_0^{\infty} \left(\int_{-\infty}^{xv} f_\xi(u) f_\zeta(v) du \right) dv, \end{aligned}$$

ezért

$$f_{\xi/\zeta}(x) = F'_{\xi/\zeta}(x) = \int_{-\infty}^0 (-v) f_\xi(xv) f_\zeta(v) dv + \int_0^{\infty} v f_\xi(xv) f_\zeta(v) dv.$$

□

2.2.3. Lemma. A $\mathbf{t}_k$ -eloszlás abszolút folytonos, és sűrűségfüggvénye

$$f_{\mathbf{t}_k}(x) = \frac{\tilde{c}_k}{\left(\frac{x^2}{k} + 1\right)^{(k+1)/2}},$$

ahol

$$\tilde{c}_k = \begin{cases} \frac{(k-1)!!}{\pi \sqrt{k} \cdot (k-2)!!}, & \text{ha } k \text{ páratlan,} \\ \frac{(k-1)!!}{2\sqrt{k} \cdot (k-2)!!}, & \text{ha } k \text{ páros.} \end{cases}$$

Speciálisan

$$f_{\mathbf{t}_1}(x) = \frac{1}{\pi(x^2 + 1)}, \quad f_{\mathbf{t}_2}(x) = \frac{1}{(x^2 + 2)^{3/2}}.$$

Tehát a $\mathbf{t}_1$ -eloszlás megegyezik a Cauchy-eloszlással.

Bizonyítás. Először meghatározzuk $\sqrt{\chi_k^2/k}$ eloszlásfüggvényét:

$$F_{\sqrt{\chi_k^2/k}}(x) = \mathbf{P}(\sqrt{\chi_k^2/k} < x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \mathbf{P}(\chi_k^2 < kx^2), & \text{ha } x > 0. \end{cases}$$

Tehát ha $x > 0$, akkor $F_{\sqrt{\chi_k^2/k}}(x) = F_{\chi_k^2}(kx^2)$, ami deriválható: $F'_{\sqrt{\chi_k^2/k}}(x) = 2kx f_{\chi_k^2}(kx^2)$, ezért $\sqrt{\chi_k^2/k}$ abszolút folytonos, és sűrűségfüggvénye

$$f_{\sqrt{\chi_k^2/k}}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ 2kx c_k (kx^2)^{k/2-1} e^{-kx^2/2}, & \text{ha } x > 0, \end{cases}$$

így ha $x > 0$, akkor $f_{\sqrt{\chi_k^2/k}}(x) = 2k^{k/2} c_k x^{k-1} e^{-kx^2/2}$. Ebből a 2.2.2. Lemma felhasználásával

$$\begin{aligned} f_{t_k}(x) &= \int_{-\infty}^{\infty} |v| f_{\eta}(xv) f_{\sqrt{\chi_k^2/k}}(v) dv = \int_0^{\infty} v \frac{1}{\sqrt{2\pi}} e^{-x^2 v^2/2} 2k^{k/2} c_k v^{k-1} e^{-kv^2/2} dv \\ &= \frac{2k^{k/2} c_k}{\sqrt{2\pi}} \int_0^{\infty} v^k e^{-(x^2+k)v^2/2} dv = \frac{2k^{k/2} c_k}{\sqrt{2\pi}} \int_0^{\infty} \frac{y^k}{(x^2+k)^{k/2}} e^{-y^2/2} \frac{dy}{(x^2+k)^{1/2}} \\ &= \frac{\tilde{c}_k}{(x^2+k)^{(k+1)/2}}, \end{aligned}$$

ahol

$$\tilde{c}_k = \frac{2k^{k/2} c_k}{\sqrt{2\pi}} \int_0^{\infty} y^k e^{-y^2/2} dy.$$

Ha $k = 1$, akkor $c_1 = 1/\sqrt{2\pi}$ és $\int_0^{\infty} y e^{-y^2/2} dy = \left[-e^{-y^2/2} \right]_{y=0}^{y=\infty} = 1$, így $\tilde{c}_1 = 1/\pi$, ezért $f_{t_1}(x) = 1/[\pi(x^2+1)]$. Ha $k = 2$, akkor $c_2 = 1/2$ és

$$\int_0^{\infty} y^2 e^{-y^2/2} dy = \frac{1}{2} \sqrt{2\pi} \int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{\sqrt{2\pi}}{2},$$

így $\tilde{c}_2 = 1$, ezért $f_{t_2}(x) = 1/(x^2+2)^{3/2}$. Továbbá $k \geq 2$ esetén parciális integrálással ($y^{n-1} \cdot y e^{-y^2/2}$ felbontással $(-e^{-y^2/2})' = y e^{-y^2/2}$ alapján)

$$\int_0^{\infty} y^k e^{-y^2/2} dy = \left[-y^{k-1} e^{-y^2/2} \right]_{y=0}^{y=\infty} + (k-1) \int_0^{\infty} y^{k-2} e^{-y^2/2} dy = (k-1) \int_0^{\infty} y^{k-2} e^{-y^2/2} dy,$$

ugyanis L'Hospital szabállyal bebizonyítható, hogy $\lim_{y \rightarrow \infty} y^{n-1} e^{-y^2/2} = 0$. Ha $k \geq 3$ páratlan, akkor

$$\int_0^{\infty} y^k e^{-y^2/2} dy = (k-1)(k-3) \dots 2 \int_0^{\infty} y e^{-y^2/2} dy = (k-1)!!,$$

amiből

$$\tilde{c}_k = \frac{2k^{k/2}}{\sqrt{2\pi}} \cdot \frac{(k-1)!!}{\sqrt{2\pi} \cdot (k-2)!!} = \frac{k^{k/2} \cdot (k-1)!!}{\pi \cdot (k-2)!!}.$$

Ha $k \geq 4$ páros, akkor

$$\int_0^\infty y^k e^{-y^2/2} dy = (k-1)(k-3) \dots 3 \int_0^\infty y^2 e^{-y^2/2} dy = (k-1)!! \sqrt{2\pi}/2,$$

amiből

$$\tilde{c}_k = \frac{2k^{k/2}}{\sqrt{2\pi}} \cdot \frac{(k-1)!! \sqrt{2\pi}/2}{2 \cdot (k-2)!!} = \frac{k^{k/2} \cdot (k-1)!!}{2 \cdot (k-2)!!}.$$

□

2.3. $\mathbf{F}_{k,\ell}$ -eloszlás

2.3.1. Definíció. Ha χ_k^2 és χ_ℓ^2 független valószínűségi változók χ_k^2 -, illetve χ_ℓ^2 -eloszlással, akkor

$$\mathbf{F}_{k,\ell} := \frac{\chi_k^2/k}{\chi_\ell^2/\ell}$$

eloszlását **$\mathbf{F}_{k,\ell}$ -eloszlásnak** nevezzük, melynek **szabadsági fokai k és ℓ** .

2.3.2. Lemma. Az $\mathbf{F}_{k,\ell}$ -eloszlás abszolút folytonos, és sűrűségfüggvénye

$$f_{\mathbf{F}_{k,\ell}}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \tilde{c}_{k,\ell} \frac{\left(\frac{k}{\ell}x\right)^{k/2-1}}{\left(\frac{k}{\ell}x + 1\right)^{(k+\ell)/2}}, & \text{ha } x > 0, \end{cases}$$

ahol

$$\tilde{c}_{k,\ell} = \begin{cases} \frac{k \cdot (k+\ell-2)!!}{\pi \ell \cdot (k-2)!!(\ell-2)!!}, & \text{ha } k \text{ és } \ell \text{ páratlan,} \\ \frac{k \cdot (k+\ell-2)!!}{2\ell \cdot (k-2)!!(\ell-2)!!}, & \text{ha } k \text{ és } \ell \text{ páros,} \\ \frac{k \cdot (k+\ell-2)!!}{4\ell \cdot (k-2)!!(\ell-2)!!}, & \text{egyébként.} \end{cases}$$

Bizonyítás. Először meghatározzuk χ_n^2/n eloszlásfüggvényét:

$$F_{\chi_n^2/n}(x) = \mathbf{P}(\chi_n^2/n < x) = \mathbf{P}(\chi_n^2 < nx) = F_{\chi_n^2}(nx),$$

ami deriválható: $F'_{\chi_n^2/n}(x) = n f_{\chi_n^2}(nx)$, ezért χ_n^2/n abszolút folytonos, és sűrűségfüggvénye

$$f_{\chi_n^2/n}(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ n c_n (nx)^{n/2-1} e^{-nx/2}, & \text{ha } x > 0. \end{cases}$$

Ebből a 2.2.2. Lemma felhasználásával

$$\begin{aligned}
f_{\mathbb{F}_{k,\ell}}(x) &= \int_{-\infty}^{\infty} |v| f_{\chi_k^2/k}(xv) f_{\chi_\ell^2/\ell}(v) \, dv = \int_0^{\infty} v c_k k^{k/2} (xv)^{k/2-1} e^{-kxv/2} c_\ell \ell^{\ell/2} v^{\ell/2-1} e^{-\ell v/2} \, dv \\
&= c_k c_\ell k^{k/2} \ell^{\ell/2} x^{k/2-1} \int_0^{\infty} v^{(k+\ell)/2-1} e^{-(kx+\ell)v/2} \, dv \\
&= c_k c_\ell k^{k/2} \ell^{\ell/2} \frac{x^{k/2-1}}{(kx+\ell)^{(k+\ell)/2}} \int_0^{\infty} y^{(k+\ell)/2-1} e^{-y} \, dy.
\end{aligned}$$

Ha $k + \ell$ páratlan, akkor parciális integrálással

$$\int_0^{\infty} y^{(k+\ell)/2-1} e^{-y} \, dy = \left(\frac{k+\ell}{2} - 1 \right) \left(\frac{k+\ell}{2} - 2 \right) \dots \frac{3}{2} \int_0^{\infty} \sqrt{y} e^{-y} \, dy,$$

ahol

$$\int_0^{\infty} \sqrt{y} e^{-y} \, dy = \int_0^{\infty} \frac{u}{\sqrt{2}} e^{-u^2/2} u \, du = \frac{\sqrt{2\pi}}{2\sqrt{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} u \, du = \frac{\sqrt{\pi}}{2}.$$

Ha $k + \ell$ páros, akkor parciális integrálással

$$\int_0^{\infty} y^{(k+\ell)/2-1} e^{-y} \, dy = \left(\frac{k+\ell}{2} - 1 \right) \left(\frac{k+\ell}{2} - 2 \right) \dots 1 \int_0^{\infty} e^{-y} \, dy,$$

ahol $\int_0^{\infty} e^{-y} \, dy = 1$. □

2.4. Normális eloszlású minta vizsgálata

2.4.1. Tétel. Ha $\xi_1, \xi_2, \dots, \xi_n$ független, $\mathcal{N}(m, \sigma^2)$ -eloszlású valószínűségi változók, akkor

- (i) $\bar{\xi}_n$ eloszlása $\mathcal{N}(m, \sigma^2/n)$,
- (ii) $\bar{\xi}_n$ és s_n^2 függetlenek,
- (iii) ns_n^2/σ^2 eloszlása χ_{n-1}^2 -eloszlás.

Ennek a tételnek a bizonyításához felhasználjuk a karakterisztikus függvényeket.

2.4.2. Definíció. A $\xi : \Omega \rightarrow \mathbb{R}$ valószínűségi változó **karakterisztikus függvénye** $\varphi_\xi : \mathbb{R} \rightarrow \mathbb{C}$,

$$\varphi_\xi(t) := \mathbb{E} e^{it\xi} := \mathbb{E} \cos(t\xi) + i \mathbb{E} \sin(t\xi).$$

A $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozó **karakterisztikus függvénye** (másképpen: a $\xi_1, \dots, \xi_k$ valószínűségi változók együttes karakterisztikus függvénye) $\varphi_\xi = \varphi_{\xi_1, \dots, \xi_k} : \mathbb{R}^k \rightarrow \mathbb{C}$,

$$\varphi_\xi(t) = \varphi_{\xi_1, \dots, \xi_k}(t_1, \dots, t_k) := \mathbb{E} e^{i\langle t, \xi \rangle} = \mathbb{E} \exp \left\{ i \sum_{j=1}^k t_j \xi_j \right\}.$$

2.4.3. Állítás. (i) Ha a $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ és az $\eta = (\eta_1, \dots, \eta_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozók függetlenek, akkor

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t), \quad t \in \mathbb{R}^k.$$

(ii) Ha a $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ és az $\eta = (\eta_1, \dots, \eta_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozók karakterisztikus függvényei megegyeznek, azaz $\varphi_\xi(t) = \varphi_\eta(t)$ bármely $t \in \mathbb{R}^k$ esetén, akkor megegyeznek az eloszlásaik, azaz $F_\xi(x) = F_\eta(x)$ bármely $x \in \mathbb{R}^k$ esetén.

(iii) A $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ és az $\eta = (\eta_1, \dots, \eta_\ell) : \Omega \rightarrow \mathbb{R}^\ell$ valószínűségi vektorváltozók akkor és csak akkor függetlenek, ha

$$\varphi_{\xi,\eta}(u, v) = \varphi_\xi(u)\varphi_\eta(v), \quad u \in \mathbb{R}^k, \quad v \in \mathbb{R}^\ell.$$

(iv) Ha $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozó, A egy $\ell \times k$ -mátrix, és $b \in \mathbb{R}^\ell$, akkor

$$\varphi_{A\xi+b}(t) = e^{i\langle t, b \rangle} \varphi_\xi(A^\top t), \quad t \in \mathbb{R}^\ell.$$

Speciálisan: ha $\xi : \Omega \rightarrow \mathbb{R}$ valószínűségi változó és $a, b \in \mathbb{R}$, akkor

$$\varphi_{a\xi+b}(t) = e^{itb} \varphi_\xi(at), \quad t \in \mathbb{R}.$$

Bizonyítás. (i). A függetlenség alapján

$$\varphi_{\xi+\eta}(t) = \mathbb{E} e^{i\langle t, \xi+\eta \rangle} = \mathbb{E} (e^{i\langle t, \xi \rangle} e^{i\langle t, \eta \rangle}) = \mathbb{E} e^{i\langle t, \xi \rangle} \mathbb{E} e^{i\langle t, \eta \rangle} = \varphi_\xi(t)\varphi_\eta(t).$$

(ii). Nehéz. (Azt lehet felhasználni, hogy zárt intervallumon értelmezett folytonos függvények egyenletesen közelíthetők trigonometrikus polinomokkal.)

(iii). Ha ξ és η függetlenek, akkor

$$\varphi_{\xi,\eta}(u, v) = \mathbb{E} e^{i(\langle u, \xi \rangle + \langle v, \eta \rangle)} = \mathbb{E} (e^{i\langle u, \xi \rangle} e^{i\langle v, \eta \rangle}) = \mathbb{E} e^{i\langle u, \xi \rangle} \mathbb{E} e^{i\langle v, \eta \rangle} = \varphi_\xi(u)\varphi_\eta(v).$$

Ha pedig teljesül

$$\varphi_{\xi,\eta}(u, v) = \varphi_\xi(u)\varphi_\eta(v), \quad u \in \mathbb{R}^k, \quad v \in \mathbb{R}^\ell,$$

akkor tekintsünk olyan $\tilde{\xi}$ és $\tilde{\eta}$ független valószínűségi vektorváltozókat, hogy ξ és $\tilde{\xi}$, valamint η és $\tilde{\eta}$ eloszlásai megegyeznek. Ekkor nyilván $\varphi_{\tilde{\xi}} = \varphi_\xi$ és $\varphi_{\tilde{\eta}} = \varphi_\eta$, valamint

$$\varphi_{\tilde{\xi}, \tilde{\eta}}(u, v) = \varphi_{\tilde{\xi}}(u)\varphi_{\tilde{\eta}}(v) \quad u \in \mathbb{R}^k, \quad v \in \mathbb{R}^\ell,$$

amiből következik, hogy

$$\varphi_{\xi,\eta}(u, v) = \varphi_{\tilde{\xi}, \tilde{\eta}}(u, v) \quad u \in \mathbb{R}^k, \quad v \in \mathbb{R}^\ell,$$

ezért a $(\xi, \eta) : \Omega \rightarrow \mathbb{R}^{k+\ell}$ és $(\tilde{\xi}, \tilde{\eta}) : \Omega \rightarrow \mathbb{R}^{k+\ell}$ valószínűségi vektorváltozók eloszlásai megegyeznek. Így a $\tilde{\xi}$ és $\tilde{\eta}$ függetlenségéből kapjuk, hogy ξ és η is függetlenek, ugyanis

$$F_{\xi,\eta}(x, y) = F_{\tilde{\xi}, \tilde{\eta}}(x, y) = F_{\tilde{\xi}}(x)F_{\tilde{\eta}}(y) = F_\xi(x)F_\eta(y), \quad x \in \mathbb{R}^k, \quad y \in \mathbb{R}^\ell.$$

(iv). Nyilván

$$\varphi_{A\xi+b}(t) = \mathbb{E}e^{i\langle t, A\xi+b \rangle} = e^{i\langle t, b \rangle} \mathbb{E}e^{i\langle A^\top t, \xi \rangle} = e^{i\langle t, b \rangle} \varphi_\xi(A^\top t),$$

ugyanis $\langle t, A\xi \rangle = t^\top(A\xi) = (A^\top t)^\top \xi = \langle A^\top t, \xi \rangle$, hiszen $u, v \in \mathbb{R}^k$ esetén $\langle u, v \rangle = u^\top v$. $\square$

Példák:

- Ha η standard normális eloszlású, akkor $\varphi_\eta(t) = e^{-t^2/2}$, ugyanis

$$\varphi_\eta(t) = \mathbb{E}e^{it\eta} = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2 - (x-it)^2/2} dx,$$

és belátható, hogy $\int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx = \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$.

Ha pedig ξ normális eloszlású (m, σ^2) paraméterekkel, akkor $\eta := (\xi - m)/\sigma$ standard normális eloszlású, tehát $\xi = \sigma\eta + m$ alapján $\varphi_\xi(t) = e^{itm} e^{-(\sigma t)^2/2} = e^{itm - \sigma^2 t^2/2}$.

- Ha ξ exponenciális eloszlású λ paraméterrel, akkor $\varphi_\xi(t) = 1/(1 - it/\lambda)$, hiszen

$$\varphi_\xi(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(it-\lambda)x} dx = \frac{\lambda}{(it - \lambda)} [e^{(it-\lambda)x}]_{x=0}^{x=\infty} = -\frac{\lambda}{it - \lambda}.$$

- Mivel a χ_2^2 -eloszlás megegyezik az $1/2$ -paraméterű exponenciális eloszlással, így annak a karakterisztikus függvénye $\varphi_{\chi_2^2}(t) = 1/(1 - 2it)$. A χ_2^n -eloszlás definíciója alapján

$$\varphi_{\chi_2^n}(t) = \varphi_{\sum_{k=1}^n \eta_k^2}(t) = \prod_{k=1}^n \varphi_{\eta_k^2}(t) = \left(\varphi_{\eta_1^2}(t) \right)^n,$$

ezért $1/(1 - 2it) = \varphi_{\chi_2^2}(t) = \left(\varphi_{\eta_1^2}(t) \right)^2$ alapján $\varphi_{\eta_1^2}(t) = 1/\sqrt{1 - 2it}$, így végülis $\varphi_{\chi_2^n}(t) = (1 - 2it)^{-n/2}$.

2.4.4. Definíció. Az $\eta = (\eta_1, \dots, \eta_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozót ***k*-dimenziós standard normális eloszlásúnak** nevezzük, ha $\eta_1, \dots, \eta_k$ független, standard normális eloszlásúak.

2.4.5. Definíció. A $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozót ***k*-dimenziós normális eloszlásúnak** nevezzük, ha létezik olyan A $k \times k$ -mátrix és $b \in \mathbb{R}^k$ vektor úgy, hogy ξ és $A\eta + b$ eloszlása megegyezik, ahol $\eta = (\eta_1, \dots, \eta_k) : \Omega \rightarrow \mathbb{R}^k$ *k*-dimenziós standard normális eloszlású.

2.4.6. Tétel. A $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozó akkor és csak akkor normális eloszlású, ha karakterisztikus függvénye

$$\varphi_\xi(t) = e^{i\langle m, t \rangle - \langle Dt, t \rangle/2}, \quad t \in \mathbb{R}^k$$

alakú, ahol $m \in \mathbb{R}^k$, D pedig egy valós, szimmetrikus, pozitív szemidefinit $k \times k$ -mátrix (azaz tetszőleges $t = (t_1, \dots, t_k) \in \mathbb{R}^k$ esetén $\langle Dt, t \rangle = \sum_{j=1}^k \sum_{\ell=1}^k d_{j,\ell} t_j t_\ell \geq 0$). Továbbá $\mathbb{E}\xi = (\mathbb{E}\xi_1, \dots, \mathbb{E}\xi_k) = m$ és $\text{cov}(\xi) = \mathbb{E}[(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^\top] = D$.

Ha D invertálható, akkor ξ abszolút folytonos, és sűrűségfüggvénye

$$f_\xi(x) = \frac{1}{(2\pi)^{k/2} \sqrt{\det(D)}} e^{-\langle D^{-1}(x-m), x-m \rangle / 2}.$$

2.4.7. Tétel. Ha $(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) : \Omega \rightarrow \mathbb{R}^{m+n}$ normális eloszlású, és $\text{cov}(\xi_k, \eta_\ell) = 0$ teljesül minden $k = 1, \dots, m$ és $\ell = 1, \dots, n$ esetén, akkor $\xi := (\xi_1, \dots, \xi_m)$ és $\eta := (\eta_1, \dots, \eta_n)$ függetlenek.

Bizonyítás. A $(\xi, \eta) := (\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n)$ várható érték vektora és kovarianciamátrixa

$$\begin{pmatrix} \mathbb{E}_\xi \\ \mathbb{E}_\eta \end{pmatrix}, \quad \begin{pmatrix} D_\xi & 0 \\ 0 & D_\eta \end{pmatrix},$$

ahol $D_\xi := \text{cov}(\xi)$, $D_\eta := \text{cov}(\eta)$. Ezért a karakterisztikus függvénye

$$\begin{aligned} \varphi_{\xi, \eta}(u, v) &= \exp \left\{ i \left\langle \begin{pmatrix} \mathbb{E}_\xi \\ \mathbb{E}_\eta \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle - \frac{1}{2} \left\langle \begin{pmatrix} D_\xi & 0 \\ 0 & D_\eta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle \right\} \\ &= \exp \left\{ i(\langle \mathbb{E}_\xi, u \rangle + \langle \mathbb{E}_\eta, v \rangle) - \frac{1}{2}(\langle D_\xi u, u \rangle + \langle D_\eta v, v \rangle) \right\} = \varphi_\xi(u) \varphi_\eta(v) \end{aligned}$$

tetszőleges $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ esetén. Ebből pedig következik a 2.4.3. Állítás (iii) pontja alapján ξ és η függetlensége. $\square$

2.4.8. Tétel. A $\xi = (\xi_1, \dots, \xi_k) : \Omega \rightarrow \mathbb{R}^k$ valószínűségi vektorváltozó akkor és csak akkor normális eloszlású, ha tetszőleges $c_1, \dots, c_k \in \mathbb{R}$ esetén $\sum_{j=1}^k c_j \xi_j$ (egydimenziós) normális eloszlású.

Bizonyítás. Ha $\xi = (\xi_1, \dots, \xi_k)$ normális eloszlású, akkor tetszőleges $c = (c_1, \dots, c_k) \in \mathbb{R}^k$ esetén $\eta := \langle c, \xi \rangle = \sum_{j=1}^k c_j \xi_j$ karakterisztikus függvénye

$$\varphi_\eta(t) = \mathbb{E} e^{it\eta} = \mathbb{E} e^{it\langle c, \xi \rangle} = \mathbb{E} e^{i\langle tc, \xi \rangle} = \varphi_\xi(tc) = e^{i\langle m, tc \rangle - \langle D(tc), tc \rangle / 2} = e^{i\langle m, c \rangle t - \langle Dc, c \rangle t^2 / 2},$$

tehát η valóban normális eloszlású $(\langle m, c \rangle, \langle Dc, c \rangle)$ paraméterekkel.

Ha pedig tetszőleges $c_1, \dots, c_k \in \mathbb{R}$ esetén $\eta := \langle c, \xi \rangle = c^\top \xi$ normális eloszlású, akkor nyilván a paraméterei $\mathbb{E}\eta = c^\top \mathbb{E}\xi$ és

$$\begin{aligned} D^2 \eta &= \mathbb{E} [(\eta - \mathbb{E}\eta)(\eta - \mathbb{E}\eta)^\top] = \mathbb{E} [c^\top (\xi - \mathbb{E}\xi) (c^\top (\xi - \mathbb{E}\xi))^\top] \\ &= c^\top \mathbb{E} [(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^\top] c = c^\top \text{cov}(\xi) c. \end{aligned}$$

Így $\xi = (\xi_1, \dots, \xi_k)$ karakterisztikus függvénye

$$\varphi_\xi(t) = \mathbb{E} e^{it\langle \xi, t \rangle} = \varphi_{\langle t, \xi \rangle}(1) = e^{it^\top \mathbb{E}\xi - t^\top \text{cov}(\xi) t / 2} = e^{i\langle \mathbb{E}\xi, t \rangle - \langle \text{cov}(\xi) t, t \rangle / 2}$$

tetszőleges $t \in \mathbb{R}^k$ esetén, tehát ξ valóban normális eloszlású. $\square$

A 2.4.1. Tétel bizonyítása. (i). Nyilván $\xi := (\xi_1, \dots, \xi_n)$ normális eloszlású, hiszen $\eta_k := (\xi_k - m)/\sigma$, $k = 1, \dots, n$ független, standard normális eloszlásúak, és $\xi_k = \sigma\eta_k + m$, $k = 1, \dots, n$ alapján a ξ vektor előállítható, mint az $\eta := (\eta_1, \dots, \eta_n)$ standard normális eloszlású vektor lineáris transzformáltja. Továbbá $\bar{\xi}_n$ a $\xi_1, \dots, \xi_n$ lineáris kombinációja, ezért a 2.4.8. Tétel alapján $\bar{\xi}_n$ valóban normális eloszlású. Továbbá nyilván $E\bar{\xi}_n = \sum_{j=1}^n E\xi_j/n = nm/n = m$ és $D^2\bar{\xi}_n = \sum_{j=1}^n D^2\xi_j/n^2 = n\sigma^2/n^2 = \sigma^2/n$.

(ii). Először a 2.4.8. Tétel alapján belátható, hogy a

$$(\xi_1 - \bar{\xi}_n, \xi_2 - \bar{\xi}_n, \dots, \xi_n - \bar{\xi}_n, \bar{\xi}_n)$$

vektor normális eloszlású, hiszen a koordinátáinak lineáris kombinációi egyúttal a normális eloszlású $(\xi_1, \xi_2, \dots, \xi_n)$ vektor koordinátáiból képezett lineáris kombinációknak is tekinthetők.

Továbbá minden $k = 1, \dots, n$ esetén teljesül $\text{cov}(\xi_k - \bar{\xi}_n, \bar{\xi}_n) = 0$, hiszen

$$\begin{aligned} \text{cov}(\xi_k - \bar{\xi}_n, \bar{\xi}_n) &= \text{cov}(\xi_k, \bar{\xi}_n) - \text{cov}(\bar{\xi}_n, \bar{\xi}_n) = \text{cov}\left(\xi_k, \frac{1}{n} \sum_{i=1}^n \xi_i\right) - \text{cov}\left(\frac{1}{n} \sum_{i=1}^n \xi_i, \frac{1}{n} \sum_{j=1}^n \xi_j\right) \\ &= \frac{1}{n} \sum_{i=1}^n \text{cov}(\xi_k, \xi_i) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(\xi_i, \xi_j) = \frac{\sigma^2}{n} - \frac{n\sigma^2}{n^2} = 0, \end{aligned}$$

ugyanis a $\xi_1, \xi_2, \dots, \xi_n$ változók függetlensége miatt

$$\text{cov}(\xi_i, \xi_j) = \begin{cases} \sigma^2, & \text{ha } i = j, \\ 0, & \text{ha } i \neq j. \end{cases}$$

Ezért a 2.4.7. Tétel alapján $(\xi_1 - \bar{\xi}_n, \xi_2 - \bar{\xi}_n, \dots, \xi_n - \bar{\xi}_n)$ és $\bar{\xi}_n$ függetlenek, amiből már következik

$$s_n^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi}_n)^2$$

és $\bar{\xi}_n$ függetlensége.

(iii). A Steiner-formula alkalmazásával

$$\frac{n}{\sigma^2} s_n^2 = \frac{1}{\sigma^2} \sum_{k=1}^n (\xi_k - \bar{\xi}_n)^2 = \sum_{k=1}^n \left(\frac{\xi_k - m}{\sigma} \right)^2 - \frac{1}{n} \left(\sum_{k=1}^n \frac{\xi_k - m}{\sigma} \right)^2 = \sum_{k=1}^n \eta_k^2 - \eta^2,$$

ahol $\eta_k := (\xi_k - m)/\sigma$, $k = 1, \dots, n$ független, standard normális eloszlásúak, és

$$\eta := \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\xi_k - m}{\sigma} = \frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}}$$

is standard normális eloszlású. Tehát

$$\frac{n}{\sigma^2} s_n^2 + \eta^2 = \sum_{k=1}^n \eta_k^2,$$

ahol a baloldalon álló összeadandók (ii) alapján függetlenek. Ezért a baloldal karakterisztikus függvénye egyrészt

$$\varphi_{ns_n^2/\sigma^2+\eta^2}(t) = \varphi_{ns_n^2/\sigma^2}(t)\varphi_{\eta^2}(t),$$

másrészt pedig

$$\varphi_{ns_n^2/\sigma^2+\eta^2}(t) = \varphi_{\sum_{k=1}^n \eta_k^2}(t) = \prod_{k=1}^n \varphi_{\eta_k^2}(t) = (\varphi_{\eta^2}(t))^n.$$

Mivel tetszőleges $t \in \mathbb{R}$ esetén $\varphi_{\eta^2}(t) = 1/\sqrt{1-2it} \neq 0$, ezért lehet vele osztani, így

$$\varphi_{ns_n^2/\sigma^2}(t) = (\varphi_{\eta^2}(t))^{n-1} = \prod_{k=1}^{n-1} \varphi_{\eta_k^2}(t) = \varphi_{\sum_{k=1}^{n-1} \eta_k^2}(t),$$

tehát ns_n^2/σ^2 valóban χ_{n-1}^2 -eloszlású. □

3. Becsléselmélet

Feladat: a minta alapeloszlásához rendelt valamely mennyiség becslése, azaz közelítése a mintaelemek felhasználásával, vagyis statisztikákkal.

3.1. A várható érték és a szórásnégyzet becslése

Jelölje az alapeloszlás (vagyis a $\xi_1, \xi_2, \dots, \xi_n$ mintaelemek közös eloszlásának) várható értékét m , szórásnégyzetét σ^2 . Nyilván

$$\mathbb{E}\bar{\xi}_n = \frac{1}{n}(\mathbb{E}\xi_1 + \dots + \mathbb{E}\xi_n) = m,$$

vagyis a mintaátlag várható értéke megegyezik az alapeloszlás várható értékével. Ezért azt mondjuk, hogy a mintaátlag **torzítatlan becslése** a várható értéknek.

Továbbá

$$\mathbb{D}^2\bar{\xi}_n = \frac{1}{n^2}(\mathbb{D}^2\xi_1 + \dots + \mathbb{D}^2\xi_n) = \frac{\sigma^2}{n} \rightarrow 0 \quad \text{ha } n \rightarrow \infty.$$

A nagy számok erős törvénye értelmében $\mathbb{P}(\lim_{n \rightarrow \infty} \bar{\xi}_n = m) = 1$, amit úgy fejezünk ki, hogy a mintaátlag **erősen konzisztens becslése** a várható értéknek.

A centrális határeloszlástétel alapján $\sqrt{n}(\bar{\xi}_n - m)/\sigma \rightarrow \mathcal{N}(0, 1)$ eloszlásban ha $n \rightarrow \infty$, hiszen

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} < x\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}^2S_n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{u^2/2} du,$$

ahol $S_n = \xi_1 + \dots + \xi_n$, $\mathbb{E}S_n = nm$, $\mathbb{D}^2S_n = n\sigma^2$.

A tapasztalati szórásnégyzet várható értéke

$$\mathbb{E}s_n^2 = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(\xi_k - \bar{\xi}_n)^2 = \frac{1}{n} \sum_{k=1}^n \mathbb{D}^2(\xi_k - \bar{\xi}_n) = \frac{n-1}{n} \sigma^2,$$

hiszen

$$\mathbb{D}^2(\xi_k - \bar{\xi}_n) = \mathbb{D}^2 \left[\left(1 - \frac{1}{n}\right) \xi_k + \frac{1}{n} \sum_{\{j:j \neq k\}} \xi_j \right] = \frac{(n-1)^2}{n^2} \mathbb{D}^2 \xi_k + \frac{1}{n^2} \sum_{\{j:j \neq k\}} \mathbb{D}^2 \xi_j = \frac{n-1}{n} \sigma^2.$$

Ezért a tapasztalati szórásnégyzet nem torítatlan becslése a szórásnégyzetnek. Viszont a **korrigált tapasztalati szórásnégyzet**:

$$s_n^{*2} := \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2 = \frac{n}{n-1} s_n^2$$

nyilván torítatlan becslése a szórásnégyzetnek: $\mathbb{E}s_n^{*2} = \sigma^2$. Be lehet látni, hogy

$$\mathbb{D}^2(s_n^{*2}) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right),$$

ahol $\mu_4 := \mathbb{E}[(\xi_1 - m)^4]$ az alapeloszlás negyedik centrális momentuma. Azt is be lehet látni, hogy a korrigált tapasztalati szórásnégyzet erősen konzisztens becslése a szórásnégyzetnek: $\mathbb{P} \left(\lim_{n \rightarrow \infty} s_n^{*2} = \sigma^2 \right) = 1$.

3.2. Pontbecslések

Feltesszük, hogy az alapeloszlás, melyből a minta származik, valamely paraméteresen megadott $\{F^{(\gamma)} : \gamma \in \Gamma\}$ eloszláscsaládból való, ahol $\Gamma \subset \mathbb{R}^k$. Feladat: az ismeretlen γ paraméter becslése.

A **maximum likelihood módszer** szerint azzal a paraméterértékkel közelítünk, mely esetén a legvalószínűbb az az esemény, hogy éppen az adott mintaelemeket kapjuk.

Példák:

- (i) Tegyük fel, hogy az alapeloszlás $\lambda > 0$ paraméterű Poisson-eloszlás, ahol λ ismeretlen paraméter, azaz $\xi_1, \xi_2, \dots, \xi_n$ független valószínűségi változók, és

$$\mathbb{P}_\lambda(\xi_i = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

és a feladat: λ becslése a $\xi_1, \xi_2, \dots, \xi_n$ minta alapján. Jelölje $k_1, k_2, \dots, k_n$ a megfigyelt értékeket. A λ paraméter $\hat{\lambda}_n$ maximum likelihood becslése olyan λ szám, ahol az

$$L_n(k_1, \dots, k_n; \lambda) := \mathbb{P}_\lambda(\xi_1 = k_1, \dots, \xi_n = k_n)$$

likelihood függvénynek globális maximumhelye van. Nyilván

$$L_n(k_1, \dots, k_n; \lambda) = \prod_{i=1}^n P_\lambda(\xi_i = k_i) = \prod_{i=1}^n \left(\frac{\lambda^{k_i}}{k_i!} e^{-\lambda} \right) = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!},$$

és a logaritmusfüggvény monotonitása miatt elegendő megkeresni a **log-likelihood függvény**:

$$\log L_n(k_1, \dots, k_n; \lambda) = -n\lambda + \sum_{i=1}^n k_i \log \lambda - \sum_{i=1}^n \log(k_i!)$$

globális maximumhelyét. Mivel a

$$\frac{\partial \log L_n(k_1, \dots, k_n; \lambda)}{\partial \lambda} = -n + \sum_{i=1}^n \frac{k_i}{\lambda} = 0$$

egyenlet egyetlen megoldása $\lambda = \sum_{i=1}^n k_i / n$, és

$$\frac{\partial^2 \log L_n(k_1, \dots, k_n; \lambda)}{\partial \lambda^2} = -\sum_{i=1}^n \frac{k_i}{\lambda^2} < 0$$

ha a $k_1, k_2, \dots, k_n$ megfigyelt értékek nem mindegyike 0, így ekkor a λ paraméter maximum likelihood becslése $\hat{\lambda}_n = \bar{\xi}_n$.

- (ii) Legyen az alapeloszlás $\lambda > 0$ paraméterű exponenciális eloszlás, ahol λ ismeretlen paraméter, azaz $\xi_1, \xi_2, \dots, \xi_n$ független valószínűségi változók melyeknek a sűrűségfüggvénye

$$f_{\xi_i}^{(\lambda)}(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ha } x > 0, \\ 0, & \text{ha } x \leq 0, \end{cases}$$

és a feladat: λ becslése a $\xi_1, \xi_2, \dots, \xi_n$ minta alapján. Jelölje $x_1, x_2, \dots, x_n$ a megfigyelt értékeket. A λ paraméter $\hat{\lambda}_n$ maximum likelihood becslése olyan x szám, ahol az

$$L_n(x_1, \dots, x_n; \lambda) := f_{\xi_1, \dots, \xi_n}^{(\lambda)}(x_1, \dots, x_n)$$

likelihood függvénynek globális maximumhelye van. Nyilván

$$L_n(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n f_{\xi_i}^{(\lambda)}(x_i) = \begin{cases} \lambda^n e^{-\lambda(x_1 + \dots + x_n)}, & \text{ha } x_1 > 0, \dots, x_n > 0, \\ 0, & \text{egyébként,} \end{cases}$$

és a logaritmusfüggvény monotonitása miatt elegendő megkeresni a **log-likelihood függvény**:

$$\log L_n(x_1, \dots, x_n; \lambda) = n \log \lambda - \lambda \sum_{i=1}^n x_i$$

globális maximumhelyét. Mivel a

$$\frac{\partial \log L_n(x_1, \dots, x_n; \lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0$$

egyenlet egyetlen megoldása $\lambda = n / \sum_{i=1}^n x_i$, és

$$\frac{\partial^2 \log L_n(k_1, \dots, k_n; \lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0,$$

ezért a λ paraméter maximum likelihood becslése $\hat{\lambda}_n = 1/\bar{\xi}_n$.

- (iii) Legyen az alapeloszlás egyenletes a $[0, \alpha]$ intervallumon, ahol $\alpha > 0$ az ismeretlen paraméter, azaz $\xi_1, \xi_2, \dots, \xi_n$ független valószínűségi változók melyeknek a sűrűségfüggvénye

$$f_{\xi_i}^{(\alpha)}(x) = \begin{cases} 1/\alpha, & \text{ha } x \in [0, \alpha], \\ 0, & \text{egyébként,} \end{cases}$$

és a feladat: α becslése a $\xi_1, \xi_2, \dots, \xi_n$ minta alapján. Az α paraméter $\hat{\alpha}_n$ maximum likelihood becslése olyan x szám, ahol az

$$L_n(x_1, \dots, x_n; \alpha) := f_{\xi_1, \dots, \xi_n}^{(\alpha)}(x_1, \dots, x_n)$$

likelihood függvénynek globális maximumhelye van. Nyilván

$$L_n(x_1, \dots, x_n; \alpha) = \prod_{i=1}^n f_{\xi_i}^{(\alpha)}(x_i) = \begin{cases} 1/\alpha^n, & \text{ha } x_1 \in [0, \alpha], \dots, x_n \in [0, \alpha], \\ 0, & \text{egyébként,} \end{cases}$$

és $x_1 \in [0, \alpha], \dots, x_n \in [0, \alpha]$ azzal ekvivalens, hogy $\alpha \geq \max_{1 \leq i \leq n} x_i$ ezért a globális maximumhely $\alpha = \max_{1 \leq i \leq n} x_i$, így az α paraméter maximum likelihood becslése $\hat{\alpha}_n = \max_{1 \leq i \leq n} \xi_i = \xi_{n:n}^*$.

3.3. Intervallumbecslések

Megint feltesszük, hogy az alapeloszlás, melyből a minta származik, valamely paraméteresen megadott $\{F^{(\gamma)} : \gamma \in \Gamma\}$ eloszláscsaládból való, ahol $\Gamma \subset \mathbb{R}^k$. Feladat: előre megadott $\varepsilon \in (0, 1)$ számhoz olyan $S(\xi_1, \dots, \xi_n)$ és $T(\xi_1, \dots, \xi_n)$ statisztikákat találni, hogy

$$P_\gamma(S(\xi_1, \dots, \xi_n) \leq \gamma \leq T(\xi_1, \dots, \xi_n)) = 1 - \varepsilon$$

teljesüljön. Ekkor azt mondjuk, hogy $[S(\xi_1, \dots, \xi_n), T(\xi_1, \dots, \xi_n)]$ egy **$1 - \varepsilon$ szintű konfidencia-intervallum**.

3.3.1. Konfidencia-intervallum a normális eloszlás várható értékére, ha ismert a szórásnégyzet

Tegyük fel, hogy a minta $\mathcal{N}(m, \sigma^2)$ -eloszlásból származik. Ebben az esetben a 2.4.1. Tétel alapján $\bar{\xi}_n \sim \mathcal{N}(m, \sigma^2/n)$, ezért

$$\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Válasszuk meg az $u_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy ha $\eta \sim \mathcal{N}(0, 1)$, akkor $P(\eta > u_{\varepsilon/2}) = \varepsilon/2$, azaz $P(\eta < u_{\varepsilon/2}) = 1 - \varepsilon/2$, vagyis $\Phi(u_{\varepsilon/2}) = 1 - \varepsilon/2$, ahol

$$\Phi(x) := F_\eta(x) = P(\eta < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

Ekkor persze a szimmetria miatt $P(\eta < -u_{\varepsilon/2}) = \varepsilon/2$, ezért

$$P(\eta \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]) = 1 - P(\eta > u_{\varepsilon/2}) - P(\eta < -u_{\varepsilon/2}) = 1 - \varepsilon$$

így

$$P_m \left(\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} \in [-u_{\varepsilon/2}, u_{\varepsilon/2}] \right) = 1 - \varepsilon,$$

vagyis

$$P_m (m \in [\bar{\xi}_n - u_{\varepsilon/2}\sigma/\sqrt{n}, \bar{\xi}_n + u_{\varepsilon/2}\sigma/\sqrt{n}]) = 1 - \varepsilon.$$

Tehát $[\bar{\xi}_n - u_{\varepsilon/2}\sigma/\sqrt{n}, \bar{\xi}_n + u_{\varepsilon/2}\sigma/\sqrt{n}]$ egy $1 - \varepsilon$ szintű konfidencia-intervallum az m paraméterre.

3.3.2. Konfidencia-intervallum a normális eloszlás várható értékére, ha nem ismert a szórásnégyzet

Tegyük fel, hogy a minta $\mathcal{N}(m, \sigma^2)$ -eloszlásból származik. Ebben az esetben a 2.4.1. Tétel alapján

$$\frac{\bar{\xi}_n - m}{s_n^*/\sqrt{n}} = \frac{\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}}}{\sqrt{s_n^{*2}/\sigma^2}} \sim t_{n-1},$$

azaz t_{n-1} -eloszlású. Válasszuk most a $t_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy $P(t_{n-1} > t_{\varepsilon/2}) = \varepsilon/2$ teljesüljön. Ekkor persze a szimmetria miatt $P(t_{n-1} < -t_{\varepsilon/2}) = \varepsilon/2$, ezért

$$P(t_{n-1} \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]) = 1 - P(t_{n-1} > t_{\varepsilon/2}) - P(t_{n-1} < -t_{\varepsilon/2}) = 1 - \varepsilon,$$

így

$$P_{m, \sigma^2} \left(\frac{\bar{\xi}_n - m}{s_n^*/\sqrt{n}} \in [-t_{\varepsilon/2}, t_{\varepsilon/2}] \right) = 1 - \varepsilon,$$

vagyis

$$P_{m, \sigma^2} (m \in [\bar{\xi}_n - t_{\varepsilon/2}s_n^*/\sqrt{n}, \bar{\xi}_n + t_{\varepsilon/2}s_n^*/\sqrt{n}]) = 1 - \varepsilon.$$

Tehát $[\bar{\xi}_n - t_{\varepsilon/2}s_n^*/\sqrt{n}, \bar{\xi}_n + t_{\varepsilon/2}s_n^*/\sqrt{n}]$ egy $1 - \varepsilon$ szintű konfidencia-intervallum az m paraméterre.

3.3.3. Konfidencia-intervallum a normális eloszlás szórásnégyzetére

Tegyük fel, hogy a minta $\mathcal{N}(m, \sigma^2)$ -eloszlásból származik. Ebben az esetben a 2.4.1. Tétel alapján ns_n^2/σ^2 egy χ_{n-1}^2 -eloszlású valószínűségi változó. Válasszunk most olyan $0 < c_1 < c_2$ számokat, hogy $P(\chi_{n-1}^2 < c_1) = P(\chi_{n-1}^2 > c_2) = \varepsilon/2$ teljesüljön. Ekkor

$$P_{m, \sigma^2} (ns_n^2/\sigma^2 \in [c_1, c_2]) = 1 - \varepsilon,$$

vagyis

$$P_{m, \sigma^2} (\sigma^2 \in [ns_n^2/c_2, ns_n^2/c_1]) = 1 - \varepsilon.$$

Tehát $[ns_n^2/c_2, ns_n^2/c_1]$ egy $1 - \varepsilon$ szintű konfidencia-intervallum a σ^2 paraméterre.

4. Statisztikai hipotézisek vizsgálata

4.1. Egymintás u -próba

Tegyük fel, hogy $\xi_1, \xi_2, \dots, \xi_n$ független, $\mathcal{N}(m, \sigma^2)$ -eloszlású valószínűségi változók, ahol σ^2 ismert, m ismeretlen. Legyen $m_0 \in \mathbb{R}$ egy rögzített szám. Feladat: döntsük el, hogy a minta alapján elfogadható-e az a nullhipotézis (azaz feltevés), hogy $m = m_0$, vagy inkább az $m \neq m_0$ ellenhipotézis (alternatíva) fogadható el? Tehát:

$$\begin{cases} H_0 : & m = m_0 & \text{nullhipotézis,} \\ H_1 : & m \neq m_0 & \text{alternatíva.} \end{cases}$$

Ha a H_0 nullhipotézis igaz, akkor a 2.4.1. Tétel alapján

$$u := \frac{\bar{\xi}_n - m_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Legyen $\varepsilon \in (0, 1)$ rögzített. Válasszuk meg az $u_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy ha $\eta \sim \mathcal{N}(0, 1)$, akkor $P(\eta > u_{\varepsilon/2}) = \varepsilon/2$. Ekkor

$$P_{m_0} (u \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]) = 1 - \varepsilon,$$

azaz $1 - \varepsilon$ valószínűséggel az u statisztika a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallumba esik.

Döntsünk a következő módon:

- Ha az u statisztika értéke a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallumba esik, akkor **elfogadjuk** a H_0 nullhipotézist, azaz a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallum az **elfogadási tartomány**.
- Ha az u statisztika értéke nem esik bele a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallumba, akkor **elvetjük** a H_0 nullhipotézist, azaz a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallum komplementere a **kritikus tartomány**.

4.1.1. Definíció. Elsőfajú hiba: annak a valószínűsége, hogy a H_0 nullhipotézist elvetjük, pedig H_0 igaz. (Ezt szokás a próba terjedelmének is nevezni.)

Másodfajú hiba: annak a valószínűsége, hogy a H_0 nullhipotézist elfogadjuk, pedig H_0 nem igaz.

Erőfüggvény: annak a valószínűsége, hogy nem követünk el másodfajú hibát, azaz a H_0 nullhipotézist elfogadjuk, amikor az igaz.

A próbát **konzisztensnek** nevezzük, ha a másodfajú hiba 0-hoz konvergál (vagyis az erőfüggvény értéke 1-hez tart), amikor a minta elemszáma végtelenhez tart.

A fenti döntési módszer esetén az elsőfajú hiba éppen ε , hiszen ha H_0 igaz, akkor

$$P_{m_0}(\mathbf{u} \notin [-u_{\varepsilon/2}, u_{\varepsilon/2}]) = \varepsilon$$

Nyilván

$$\mathbf{u} = \frac{\bar{\xi}_n - m_0}{\sigma/\sqrt{n}} = \frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} + \frac{m - m_0}{\sigma/\sqrt{n}},$$

ahol

$$\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

ezért a másodfajú hiba valószínűsége

$$\begin{aligned} P_m(\mathbf{u} \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]) &= P_m\left(\frac{\bar{\xi}_n - m}{\sigma/\sqrt{n}} \in \left[-u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}, u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}\right]\right) \\ &= \Phi\left(u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}\right) - \Phi\left(-u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}\right), \end{aligned}$$

ahol Φ a standard normális eloszlás eloszlásfüggvénye, azaz

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-v^2/2} dv.$$

Ezért az erőfüggvény:

$$E_n(m) = 1 - \Phi\left(u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-u_{\varepsilon/2} - \frac{m - m_0}{\sigma/\sqrt{n}}\right).$$

A próba konzisztens, azaz rögzített m esetén $\lim_{n \rightarrow \infty} E_n(m) = 1$ (ezért $n \rightarrow \infty$ esetén a másodfajú hiba 0-hoz konvergál), ugyanis $\lim_{x \rightarrow \infty} \Phi(x) = 1$, $\lim_{x \rightarrow -\infty} \Phi(x) = 0$. Továbbá rögzített n mintaelemszám esetén $\lim_{m \rightarrow \infty} E_n(m) = 1$ és $\lim_{m \rightarrow -\infty} E_n(m) = 0$ (tehát $m \rightarrow \infty$ vagy $m \rightarrow -\infty$ esetén a másodfajú hiba 0-hoz konvergál). Az E_n függvénynek minimuma van az m_0 pontban, melynek értéke ε , ezért a másodfajú hiba nagy lesz (mégpedig $1 - \varepsilon$ közeli), ha m közel van m_0 -hoz.

Nyilván ε csökkentése esetén az elsőfajú hiba valószínűsége csökken, viszont $u_{\varepsilon/2}$ növekszik, ezért a másodfajú hiba növekszik. (Tehát a kétféle hiba valószínűsége ellentétes irányban mozog!)

4.2. Egymintás t-próba

Tegyük fel, hogy $\xi_1, \xi_2, \dots, \xi_n$ független, $\mathcal{N}(m, \sigma^2)$ -eloszlású valószínűségi változók, ahol σ^2 és m is ismeretlenek. Legyen $m_0 \in \mathbb{R}$ egy rögzített szám. Legyen

$$\begin{cases} H_0 : & m = m_0 & \text{nullhipotézis,} \\ H_1 : & m \neq m_0 & \text{alternatíva.} \end{cases}$$

Tekintsük a

$$t := \frac{\bar{\xi}_n - m_0}{\sqrt{s_n^{*2}/n}}$$

statisztikát. Ha a H_0 nullhipotézis igaz, akkor a 2.4.1. Tétel alapján a

$$\frac{\bar{\xi}_n - m_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1), \quad \frac{n}{\sigma^2} s_n^2 = \frac{n-1}{\sigma^2} s_n^{*2} \sim \chi_{n-1}^2$$

statisztikák függetlenek, ezért

$$t = \frac{\frac{\bar{\xi}_n - m_0}{\sigma/\sqrt{n}}}{\sqrt{\left(\frac{n-1}{\sigma^2} s_n^{*2}\right) / (n-1)}} \sim t_{n-1},$$

azaz t_{n-1} -eloszlású. Legyen $\varepsilon \in (0, 1)$ rögzített. Válasszuk meg a $t_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy $P(t_{n-1} > t_{\varepsilon/2}) = \varepsilon/2$. Ekkor

$$P_{m_0}(t \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]) = 1 - \varepsilon,$$

azaz $1 - \varepsilon$ valószínűséggel a t statisztika a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallumba esik.

Döntsünk a következő módon:

- Ha a t statisztika értéke a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallumba esik, akkor elfogadjuk a H_0 nullhipotézist, azaz a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallum az elfogadási tartomány.
- Ha az t statisztika értéke nem esik bele a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallumba, akkor elvetjük a H_0 nullhipotézist, azaz a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallum komplementere a kritikus tartomány.

Nyilván az elsőfajú hiba éppen ε , hiszen ha H_0 igaz, akkor

$$P_{m_0}(t \notin [-t_{\varepsilon/2}, t_{\varepsilon/2}]) = \varepsilon.$$

4.3. Kétmintás u-próba

Tegyük fel, hogy a $\xi_1, \xi_2, \dots, \xi_k$ $\mathcal{N}(m_\xi, \sigma_\xi^2)$ -eloszlású, és az $\eta_1, \eta_2, \dots, \eta_\ell$ $\mathcal{N}(m_\eta, \sigma_\eta^2)$ -eloszlású valószínűségi változók függetlenek, ahol σ_ξ^2 és σ_η^2 ismertek, m_ξ és m_η ismeretlenek. Legyen

$$\begin{cases} H_0 : & m_\xi = m_\eta & \text{nullhipotézis,} \\ H_1 : & m_\xi \neq m_\eta & \text{alternatíva.} \end{cases}$$

Ha a H_0 nullhipotézis igaz, akkor a 2.4.1. Tétel alapján

$$u := \frac{\bar{\xi}_k - \bar{\eta}_\ell}{\sqrt{\frac{\sigma_\xi^2}{k} + \frac{\sigma_\eta^2}{\ell}}} \sim \mathcal{N}(0, 1).$$

Legyen $\varepsilon \in (0, 1)$ rögzített. Válasszuk meg az $u_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy ha $\eta \sim \mathcal{N}(0, 1)$, akkor $P(\eta > u_{\varepsilon/2}) = \varepsilon/2$. Ekkor megint vehetjük elfogadási tartománynak a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallumot, kritikus tartománynak pedig ennek a komplementerét.

4.4. Kétmintás t-próba

Tegyük fel, hogy a $\xi_1, \xi_2, \dots, \xi_k \sim \mathcal{N}(m_\xi, \sigma_\xi^2)$ -eloszlású, és az $\eta_1, \eta_2, \dots, \eta_\ell \sim \mathcal{N}(m_\eta, \sigma_\eta^2)$ -eloszlású valószínűségi változók függetlenek, ahol $\sigma_\xi^2, \sigma_\eta^2, m_\xi$ és m_η is ismeretlenek. Legyen

$$\begin{cases} H_0 : m_\xi = m_\eta & \text{nullhipotézis,} \\ H_1 : m_\xi \neq m_\eta & \text{alternatíva.} \end{cases}$$

Ha a H_0 nullhipotézis igaz, akkor a 2.4.1. Tétel alapján

$$t := \frac{\bar{\xi}_k - \bar{\eta}_\ell}{\sqrt{\left(\frac{1}{k} + \frac{1}{\ell}\right) \frac{(k-1)s_{\xi,k}^{*2} + (\ell-1)s_{\eta,\ell}^{*2}}{k+\ell-2}}} \sim t_{k+\ell-2}.$$

Legyen $\varepsilon \in (0, 1)$ rögzített. Válasszuk meg a $t_{\varepsilon/2} \in (0, \infty)$ számot úgy, hogy $P(t_{k+\ell-2} > t_{\varepsilon/2}) = \varepsilon/2$. Ekkor megint vehetjük elfogadási tartománynak a $[-t_{\varepsilon/2}, t_{\varepsilon/2}]$ intervallumot, kritikus tartománynak pedig ennek a komplementerét.

4.5. F-próba

Tegyük fel, hogy a $\xi_1, \xi_2, \dots, \xi_k \sim \mathcal{N}(m_\xi, \sigma_\xi^2)$ -eloszlású, és az $\eta_1, \eta_2, \dots, \eta_\ell \sim \mathcal{N}(m_\eta, \sigma_\eta^2)$ -eloszlású valószínűségi változók függetlenek, ahol $\sigma_\xi^2, \sigma_\eta^2, m_\xi$ és m_η is ismeretlenek. Legyen

$$\begin{cases} H_0 : \sigma_\xi^2 = \sigma_\eta^2 & \text{nullhipotézis,} \\ H_1 : \sigma_\xi^2 \neq \sigma_\eta^2 & \text{alternatíva.} \end{cases}$$

Tekintsük az

$$F := \frac{s_{\eta,\ell}^{*2}}{s_{\xi,k}^{*2}}$$

statisztikát. Ha a H_0 nullhipotézis igaz, akkor a 2.4.1. Tétel alapján

$$F = \frac{\left(\frac{\ell-1}{\sigma_\eta^2} s_{\eta,\ell}^{*2}\right) / (\ell-1)}{\left(\frac{k-1}{\sigma_\xi^2} s_{\xi,k}^{*2}\right) / (k-1)} \sim F_{\ell-1, k-1},$$

azaz $F_{\ell-1,k-1}$ -eloszlású. Legyen $\varepsilon \in (0, 1)$ rögzített. Válasszunk olyan $0 < c_1 < c_2$ számokat, hogy $P(F_{\ell-1,k-1} < c_1) = P(F_{\ell-1,k-1} > c_2) = \varepsilon/2$. Ekkor vehetjük elfogadási tartománynak a $[c_1, c_2]$ intervallumot, kritikus tartománynak pedig ennek a komplementerét.